УДК 519.24 DOI: 10.20998/2079-0023.2020.01.02

*А. А. ПАВЛОВ, М. Н. ГОЛОВЧЕНКО*

# ПОСТРОЕНИЕ ОДНОМЕРНОЙ И МНОГОМЕРНОЙ ПОЛИНОМИАЛЬНОЙ РЕГРЕССИИ ПО ИЗБЫТОЧНОМУ ОПИСАНИЮ С ИСПОЛЬЗОВАНИЕМ АКТИВНОГО ЭКСПЕРИМЕНТА

Рассматривается задача построения многомерной полиномиальной регрессии по заданному ее избыточному описанию на основе результатов активного эксперимента. Избыточное описание означает включение в него членов, возможно, отсутствующих в структуре исследуемой регрессии. Таким образом, возникает проблема по результатам активного эксперимента не только оценить значения неизвестных коэффициентов многомерной полиномиальной регрессии, но и исключить из ее избыточного описания лишние члены. Решение поставленной задачи базируется на: (а) получении новых свойств коэффициентов нормированных ортогональных полиномов Форсайта; (б) возможности сведения задачи оценки неизвестных коэффициентов нелинейных членов многомерной полиномиальной регрессии к задаче оценки коэффициентов множества одномерных полиномиальных регрессий и решения соответствующих систем линейных равенств; (в) использовании метода для исключения лишних членов многомерной нелинейной полиномиальной регрессии, который органически включает в себя как методологию кластерного анализа, так и основную идею метода группового учета аргументов – разбиение экспериментальных данных на два множества, одно из которых не используется для оценок неизвестных коэффициентов многомерной полиномиальной регрессии, заданной избыточным описанием.

**Ключевые слова:** многомерная полиномиальная регрессия, нормированные ортогональные полиномы Форсайта, избыточное описание, метод группового учета аргументов, кластерный анализ, линейные равенства

*О. А. ПАВЛОВ, М. М. ГОЛОВЧЕНКО*

## ПОБУДОВА ОДНОВИМІРНОЇ І БАГАТОВИМІРНОЇ ПОЛІНОМІАЛЬНОЇ РЕГРЕСІЇ ЗА НАДЛИШКОВИМ ОПИСОМ З ВИКОРИСТАННЯМ АКТИВНОГО ЕКСПЕРИМЕНТУ

Розглядається задача побудови багатовимірної поліноміальної регресії за її заданим надлишковим описом на основі результатів активного експерименту. Надлишковий опис означає включення в нього членів, можливо, відсутніх в структурі досліджуваної регресії. Таким чином, виникає проблема за результатами активного експерименту не тільки оцінити значення невідомих коефіцієнтів багатовимірної поліноміальної регресії, але і виключити з її надлишкового опису зайві члени. Розв'язання поставленої задачі базується на: (а) отриманні нових властивостей коефіцієнтів нормованих ортогональних поліномів Форсайта; (б) можливості зведення задачі оцінки невідомих коефіцієнтів нелінійних членів багатовимірної поліноміальної регресії до задачі оцінки коефіцієнтів множини одновимірних поліноміальних регресій і розв'язання відповідних систем лінійних рівностей; (в) використанні методу для виключення зайвих членів багатовимірної нелінійної поліноміальної регресії, який органічно включає в себе як методологію кластерного аналізу, так і основну ідею методу групового урахування аргументів – розбиття експериментальних даних на дві множини, одна з яких не використовується для оцінок невідомих коефіцієнтів багатовимірної поліноміальної регресії, заданої надлишковим описом.

**Ключові слова:** багатовимірна поліноміальна регресія, нормовані ортогональні поліноми Форсайта, надлишковий опис, метод групового урахування аргументів, кластерний аналіз, лінійні рівності

*A. A. PAVLOV, M. N. HOLOVCHENKO*

## UNIVARIATE AND MULTIVARIATE POLYNOMIAL REGRESSION CONSTRUCTION FROM A REDUNDANT REPRESENTATION USING AN ACTIVE EXPERIMENT

We consider the problem of a multidimensional polynomial regression construction from a given redundant representation based on the results of an active experiment. Redundant representation means inclusion in it the members which are possibly absent in the structure of the studied regression. Thus, we have a problem not only to estimate the values of the unknown coefficients of multidimensional polynomial regression from the results of an active experiment, but also to eliminate the redundant members from its redundant representation. The solution to this problem is based on: (a) obtaining new properties of the coefficients of normalized orthogonal polynomials of Forsythe; (b) possibility of reducing the problem of estimating the unknown coefficients for nonlinear members of multivariate polynomial regression to the problem of estimating the coefficients for the set of univariate polynomial regressions and solving the corresponding systems of linear equalities; (c) using the method to eliminate the redundant members of multidimensional nonlinear polynomial regression which organically includes both the methodology of cluster analysis and the main idea of the group method of data handling – dividing the experimental data into two sets, one of which is not used to estimate unknown coefficients of multidimensional polynomial regression given by a redundant representation.

**Keywords:** multi-dimensional polynomial regression, normalized orthogonal polynomials of Forsythe, redundant representation, group method of data handling, cluster analysis, linear equalities

**Введение.** Решение задачи нахождения структуры и оценки коэффициентов исследуемой многомерной полиномиальной регрессии по ее избыточному описанию с использованием активного эксперимента основано на использовании результатов, полученных в [1–10], на основе метода, излагаемого ниже, использующего основные идеи метода группового учета аргументов А. Г. Ивахненко, а также кластерного анализа.

Излагаемые ниже результаты являются естественным обобщением алгоритмов построения одномерной и многомерной полиномиальной регрессии по избыточному описанию с использованием активного эксперимента, изложенных в [1–9]. Некоторые другие алгоритмы построения регрессионных моделей и их применения описаны в [11, 12].

**Общие теоретические положения.** В [1–10] рассмотрены две задачи регрессионного анализа:

1) Построение одномерной полиномиальной регрессии по избыточному описанию с использованием активного эксперимента.

Избыточное описание задано в виде

$$Y(x) = \theta_0 + \theta_1 x + \ldots + \theta_r x^r + E\,, \qquad (1)$$

где $x$ – детерминированная переменная, к которой применим активный эксперимент. Имеется в виду, что на вход объекта, модель которого представлена в виде (1), можно подавать любые действительные значения скалярного входа $x$ и измерять выходное значение объекта $Y(x)$;

$\theta_0, \theta_1, \ldots, \theta_r$ – неизвестные коэффициенты, часть из которых, возможно, равна тождественно нулю;

$E$ – случайная величина с произвольным распределением и нулевым математическим ожиданием ($ME = 0$) и конечной дисперсией. Проведено $n$ экспериментов, в результате которых получены две выборки объема $n$ $(x_i,\ i = \overline{1,n};\ Y(x_i) = y_i,\ i = \overline{1,n})$. По результатам эксперимента необходимо исключить из (1) лишние члены и оценить оставшиеся коэффициенты $\theta_j$ $(\forall|\theta_j| > 0$ в описании (1)). Для решения этой задачи используется метод наименьших квадратов с использованием нормированных ортогональных полиномов Форсайта.

Общеизвестные результаты, изложенные в [5], позволяют получить следующие оценки $\widehat{\theta_j},\ j = \overline{0,r}$:

$$\widehat{\theta_j} = \widehat{\omega_r} q_{rj} + \ldots + \widehat{\omega_j} q_{jj},\quad j = \overline{0,r}, \tag{2}$$

где $\widehat{\omega_j}$ – оценки весовых коэффициентов, полученных методом наименьших квадратов с использованием нормированных ортогональных полиномов Форсайта;

$q_{ij}$ – коэффициенты этих полиномов.

Дисперсии оценок имеют вид [5]:

$$D(\widehat{\theta_j}) = \sigma^2 \sum_{i=r}^{j} q_{ij}^2 . \tag{3}$$

Вычислительные эксперименты, приведенные в [7–9], показывают, что с увеличением $j$ при фиксированном числе экспериментов на порядки уменьшается $D\widehat{\theta_j}$. Действительно, в проведенных экспериментах получили: при $n = 10$, $D\widehat{\theta_0} = \sigma^2 \cdot 0{,}40\ldots$, $D\widehat{\theta_1} = \sigma^2 \cdot$ $\cdot\, 0{,}0024\ldots$, $D\widehat{\theta_2} = \sigma^2 \cdot 4{,}26 \cdot 10^{-6}$, $D\widehat{\theta_3} = \sigma^2 \cdot 1{,}28 \cdot$ $\cdot\, 10^{-5}$. Предлагалось исключать из модели (1) члены, у которых оценки коэффициентов по модулю близки к нулю и исчезающе малы значения дисперсий оценок этих коэффициентов (использование закона трех сигм). Хотя использование нормированных ортогональных полиномов Форсайта облегчает решение проблемы вырожденности при использовании метода наименьших квадратов, однако, эта проблема остается, что приводит к необходимости проведения громоздких вычислений с использованием большой памяти для нахождения коэффициентов $q_{ij}$ нормированных ортогональных полиномов Форсайта.

Любые упрощения в процессе вычисления коэффициентов $\{q_{ij}\}$, гарантирующих достоверность конечного результата, упрощают решение проблемы в целом.

В [9] приведены рекомендации по повышению эффективности проведения активного эксперимента для нахождения оценок коэффициентов одномерной полиномиальной регрессии по ее избыточному описанию. Обоснованность этих рекомендаций базируется на полученных в [3, 4] следующих теоретических результатах.

1) Если $q_{jl}^x,\ \forall j = \overline{0,r},\ \forall l = \overline{0,j}$, – это коэффициенты нормированных ортогональных полиномов Форсайта, вычисленных по $x_i,\ i = \overline{1,n}$, а $q_{jl}^z,\ \forall j = \overline{0,r},\ \forall l = \overline{0,j}$ – это коэффициенты нормированных ортогональных полиномов Форсайта, вычисленных по $z_i,\ i = \overline{1,n},\ z_i = kx_i,\ i = \overline{1,n},\ k > 0$, то

$$q_{jl}^z = \frac{1}{k^j} q_{jl}^x,\ D\widehat{\theta_j^z} = \sigma^2 \sum_{l=r}^{j} \left(\frac{1}{k^j} q_{jl}^x\right) = \left(\frac{1}{k^j}\right)^2 D\widehat{\theta_j^x}.$$

2) Повторяющиеся эксперименты. Этот результат может быть использован только в случае, когда $r$ – известная степень избыточного описания одномерной полиномиальной регрессии.

Пусть на вход объекта подается повторяющаяся последовательность значений $x_1, \ldots, x_{r+p}$, $x_1, \ldots, x_{r+p}$, …, где $p > 1$. Оказывается [4], в этом случае не требуется нормированные ортогональные полиномы Форсайта строить по значениям $x_1, \ldots, x_{r+p}$, $x_1, \ldots, x_{r+p}$, … скалярного аргумента $x$, в котором последовательность чисел $x_1, \ldots, x_{r+p}$ повторяется $l$ раз. Доказано [4], что оценки коэффициентов $\widehat{\theta_j},\ j = \overline{0,r}$, не изменяются при усреднении результатов экспериментов. Пусть

$$\boldsymbol{X} = (x_1,\ \ldots, x_n),\ n = k + p;$$

$$\boldsymbol{X}' = (x_{11}, x_{21},\ \ldots, x_{l1}, x_{12}, x_{22}, \ldots, x_{l2}, \ldots, x_{1n}, x_{2n},\ \ldots, x_{ln}),$$

где $x_{ki} = x_i,\ \forall k = \overline{1,l},\ \forall i = \overline{1,n}$;

$$\boldsymbol{Y}' = (y_{11}, y_{21},\ \ldots, y_{l1}, \ldots, y_{1n}, y_{2n},\ \ldots, y_{ln});$$

$$\boldsymbol{Y} = \left(\frac{\sum_{k=1}^{l} y_{k1}}{l}, \frac{\sum_{k=1}^{l} y_{k2}}{l}, \cdots, \frac{\sum_{k=1}^{l} y_{kn}}{l}\right),$$

где $y_{ij}$ – это соответствующее значение выхода объекта в статистическом эксперименте.

Пусть $Q_j(x),\ j = \overline{0,r}$, – нормированные ортогональные полиномы Форсайта, построенные по выборке $\boldsymbol{X}$, а $Q_j'(x),\ j = \overline{0,r}$, – нормированные ортогональные полиномы Форсайта, построенные по выборке $\boldsymbol{X}'$. В [4] показано, что

$$Q_j'(x) = \frac{Q_j(x)}{\sqrt{l}},\ j = \overline{0,r},\ (x_{ki} = x_i,\ k = \overline{1,l}).$$

Из этого следует [4], что оценки $\widehat{\theta_j'} = \widehat{\theta_j},\ j = \overline{0,r}$, где $\widehat{\theta_j'}$ получены по $\boldsymbol{X}', \boldsymbol{Y}'$, а $\widehat{\theta_j}$ – по $\boldsymbol{X}, \boldsymbol{Y}$. Задача регрессии по $\boldsymbol{X}, \boldsymbol{Y}$ имеет вид

$$Y(x) = \theta_0 + \theta_1 x + \ \ldots + \theta_r x^r + E_1,$$

где $x$ принимает значения $x_1,\ \ldots, x_n,\ n = r + p,\ p\ \geq 1$, $DE_1 = \ \sigma^2/l$.

Таким образом [3, 4], варьируя параметрами нахождения наиболее эффективной области, из которой выбираются значения $x_1, \ldots, x_n$, выбрав $l$ – количество повторов серий, решив проблему нахождения практически достоверных значений коэффициентов нормированных ортогональных полиномов Форсайта, можем до проведения активного эксперимента гарантировать получение необходимых значений дисперсий $D\widehat{\theta}_j$, $j=\overline{0,r}$, на минимальном объеме данных.

**Построение многомерной полиномиальной регрессии по избыточному описанию с использованием активного эксперимента.** Пусть избыточное описание многомерной функции имеет вид [6]:

$$\overline{y}(\overline{\boldsymbol{x}})=\sum b_{i_1,\dots,i_t}^{j_1,\dots,j_t}\left(x_{i_1}\right)^{j_1}\left(x_{i_2}\right)^{j_2}\cdots\left(x_{i_t}\right)^{j_t}+E,\quad \forall(i_1,\dots,i_t)\in K, \forall(j_1,\dots,j_t)\in K(i_1,\dots,i_t) \qquad (4)$$

где $\overline{\boldsymbol{x}}=(x_1,\dots,x_n)^{\mathrm{T}}$ – детерминированный вектор входных переменных;

$b_{i_1,\dots,i_t}^{j_1,\dots,j_t}$ – неизвестные коэффициенты;

$j_i$ – натуральные числа;

$i_j$ – натуральные индексы из множеств $\{1,\dots,n\}$;

$E$ – случайная величина с нулевым математическим ожиданием и ограниченной неизвестной дисперсией.

В [1, 4] показано, что проводя активный эксперимент для случаев, когда а) $x_1=x_2=\dots=x_n=x$; б) фиксируется $x_i$, $i=\overline{1,n}$, остальные переменные $x_1=x_2==\dots=x_{i-1}=x_{i+1}=\dots=x_n=x$ изменяются одинаково, во многих случаях решая последовательные задачи построения одномерных полиномиальных регрессий и соответствующих систем линейных равенств, находят оценки всех коэффициентов многомерной регрессии при нелинейных членах, тем самым сводя исходную задачу к задаче построения многомерной линейной регрессии с использованием активного эксперимента, которая, как известно, легко решается.

**Проблема исследования.** 1) Верхние границы дисперсий оценок $\hat{b}_{i_1,\dots,i_t}^{j_1,\dots,j_t}$ (являющиеся результатом решения систем линейных неравенств, правые части которых являются оценками коэффициентов соответствующих одномерных дисперсий [1, 2, 6]) являются достаточно грубыми, и достоверно оценить, какие члены избыточного описания многомерной регрессии (4) в истинном ее выражении отсутствуют, не представляется возможным.

2) Та же проблема возникает при построении одномерной полиномиальной регрессии, если значение $D\widehat{\theta}_j$ таково, что с учетом закона трех сигм (случайная величина $E$ распределена произвольно) вероятность утверждения того, что $\theta_j=0$, существенно меньше единицы.

Сформулированная в данной статье проблема решается с помощью модификации изложенного выше метода построения одномерной и многомерной полиномиальной регрессии по ее избыточному описанию с применением активного эксперимента [1–9], которая базируется:

а) на основной идее метода группового учета аргументов А. Г. Ивахненко [10] (разбиение экспериментальных данных на множество данных, по которым находятся оценки неизвестных коэффициентов, и проверочное множество данных, в этом построении не участвующих);

б) на базовых идеях кластерного анализа.

**Модифицированный метод построения одномерной и многомерной полиномиальной регрессии** по избыточному описанию с использованием активного эксперимента.

*Примечание.* Если в избыточном описании многомерной полиномиальной регрессии переменные $x_{i_p}$, $p=\overline{1,k}$, в члены, их содержащие, входят в виде $1/x_{i_p}$, то заменой $1/x_{i_p}=z_{i_p}$, $p=\overline{1,k}$, эту задачу сводим к предыдущей.

1.1. Находим выражение для многомерной полиномиальной регрессии по ее избыточному описанию с использованием активного эксперимента, в полном соответствии с теоретическими положениями, изложенными в [1–9]. При этом предполагается, что найдены оценки коэффициентов при всех нелинейных членах регрессии, а оценки коэффициентов при переменных $x_i$, $i=\overline{1,n}$, найдены обычным методом наименьших квадратов для линейной многомерной регрессии.

1.2. Реализуем активный эксперимент в области изменения аргументов, в которой наиболее вероятно будет эксплуатироваться модель; желательно, при выполнении условия $\forall_i\,|x_i|>1$. В этом случае, нелинейные члены построенной многомерной полиномиальной регрессии не принимают малых значений (члены, у которых $\left|\hat{b}_{i_1,\dots,i_t}^{j_1,\dots,j_t}\right|\approx 0$, предварительно исключены). Множество значений аргументов и выходной величины обозначим через $\boldsymbol{XY}=\{\overline{x_l}y_l\}$, $l=\overline{1,p}$. Все члены многомерной полиномиальной регрессии (кроме линейных $b_i x_i$, $i=\overline{1,n}$) вида $\hat{b}_{i_1,\dots,i_t}^{j_1,\dots,j_t}\left(x_{i_1}\right)^{\mathrm{j}_1}\cdots\left(x_{i_t}\right)^{\mathrm{j}_t}$ переименуем $J_i$, $i=\overline{1,L}$, в порядке убывания величин $\sum_{l=1}^{p}\left|\hat{b}_{i_1,\dots,i_t}^{j_1,\dots,j_t}\left(x_{i_1 l}\right)^{j_1}\left(x_{i_2 l}\right)^{j_2}\cdots\left(x_{i_t l}\right)^{j_t}\right|$, где $x_{i_1 l},\dots,x_{i_t l}$ – значения детерминированных величин $x_{i_1},\dots,x_{i_t}$ в $l$-м эксперименте, $l=\overline{1,p}$.

2.1. Разбиваем множество $\{J_i, i=\overline{1,L}\}$ на два класса $M_1$ и $M_2$ следующим образом.

2.1.1. $J_1\in M_1$, $J_L\in M_2$. Если

$$J_1-J_2<J_2-J_L, \qquad (5)$$

то $J_2\in M_1$, в противном случае $J_2\in M_2$. Получено разбиение множества $\{J_i, i=\overline{1,L}\}$ на классы: $M_1=\{J_1\}$, $M_2=\{J_2,\dots,J_L\}$.

2.1.2. Если

$$J_{L-1}-J_L<\frac{1}{2}(J_1+J_2)-J_{L-1}, \qquad (6)$$

то $J_{L-1}\in M_2$, в противном случае $J_{L-1}\in M_1$, Получено окончательное разбиение: $M_1=\{J_1,J_2,\dots,J_{L-1}\}$, $M_2==\{J_L\}$.

2.1.3. Описанная выше процедура продолжается, если выполнены неравенства (5), (6). $(t-1)$-й шаг: $M_1=\{J_1,\dots,J_{t-1}\}$, $M_2=\{J_L,J_{L-1},\dots,J_{L-t+1}\}$. Если выполняется неравенство

$$\frac{1}{t-1}\sum_{i=1}^{t-1} J_i - J_t < J_t - \frac{1}{t-1}\sum_{j=0}^{t-1} J_{L-j}, \qquad (7)$$

то $J_{\mathbf{t}} \in M_{\mathbf{1}}$, в противном случае $J_{\mathbf{t}} \in M_{\mathbf{2}}$, Текущее разбиение множества $\{J_i, i = \overline{1,L}\}$ на два класса получено: $M_{\mathbf{1}} = \{J_i, i = \overline{1,t-1}\}$, $M_{\mathbf{2}} = \{J_{L-j}, j = \overline{0,t-1},\ J_t\}$. Если (7) выполнено, то если выполняется условие

$$J_{L-t} - \frac{1}{t-1}\sum_{j=0}^{t-1} J_{L-j} < \frac{1}{t-1}\sum_{i=1}^{t-1} J_i - J_{L-t},$$

то $J_{L-\mathbf{t}} \in M_{\mathbf{2}}$, в противном случае $J_{L-\mathbf{t}} \in M_{\mathbf{1}}$, Текущее разбиение множества $\{J_i, i = \overline{1,L}\}$ на два класса завершено: $M_{\mathbf{1}} = \{J_i, i = \overline{1,t-1}, J_{L-t}\}$, $M_{\mathbf{2}} = \{J_{L-j}, j = \overline{0,t-1}\}$.

Описанная процедура конечна и завершается текущим разбиением множества $\{J_i, i = \overline{1,L}\}$ на два класса $M_{\mathbf{1}}$ и $M_{\mathbf{2}}$.

2.2. Аналогичной процедурой множество $M_{\mathbf{2}}$ разбивается на два класса, получаем $M_{\mathbf{1}}, M_2^1, M_2^2$. Множество $M_2^2$ аналогичной процедурой разбивается на два класса $M_3^1, M_3^2$, и т. д. Эта процедура является конечной. В результате получаем окончательное разбиение множества $\{J_i, i = \overline{1,L}\}$ на классы: $M_{\mathbf{1}},\ M_2^1,\ M_3^1,\ \ldots, M_T^1$. Возможна ситуация, когда $M_T^1$ содержит один либо два элемента. В этом случае $M_T^1 = M_{T-1}^2$.

3.1. Задаем первый параметр управления вычислительным процессом $\Delta_{\mathbf{1}}$ и находим максимальное натуральное число $T_{\mathbf{1}}$, для которого выполняется

$$\min\{\forall J_i \in M_{T_1-1}^1\} - \max\{\forall J_i \in M_{T_1}^1\} \geq \Delta_{\mathbf{1}} \qquad (8)$$

($\Delta_{\mathbf{1}}$ – достаточно большое число). Если $T_{\mathbf{1}}$ найдено, то находятся две среднеквадратичные ошибки

$$\sigma^2(f_1) = \sum_{l=1}^{p}\left(y_l - f_1(\overline{x_l})\right)^2,$$
$$\sigma^2(f_2) = \sum_{l=1}^{p}\left(y_l - f_2(\overline{x_l})\right)^2,$$

где $f_1(\overline{x})$ – построенная в п. 1 полиномиальная регрессия;

$f_2(\overline{x})$ получена из $f_1(\overline{x})$ исключением из нее всех элементов из классов $M_T^1,\ \ldots, M_{T_1}^1$.

Если

$$\sigma^2(f_2) \leq \sigma^2(f_1), \qquad (9)$$

то текущее выражение для полиномиальной регрессии равно $f_2(\overline{x})$. Вводится второй параметр управления вычислительным процессом $\Delta_{\mathbf{2}}$ ($\Delta_{\mathbf{2}}$ может равняться $\Delta_{\mathbf{1}}$) и описанная выше процедура повторяется до тех пор, пока либо не выполнится условие (8), либо нарушится неравенство (9). Пусть $f_t(\overline{x})$ – это текущее выражение полиномиальной регрессии, для которого либо не выполнялось текущее условие вида (8), либо нарушилось неравенство вида (9).

3.2. Пусть $f_t(\overline{x})$ содержит члены из классов {линейные члены} $\cup \{M_{\mathbf{1}},\ M_2^1,\ \ldots, M_t^1\}$.

3.2.1. Из $f_t(\overline{x})$ исключаются члены множества $M_t^1$. Обозначим ее $f_t^1(\overline{x})$. Если выполняется

$$\sigma^2(f_t^1) \leq \sigma^2(f_t), \qquad (10)$$

то из $f_t(\overline{x})$ исключаются члены множества $M_t^1$, и она обозначается $f_{t+1}(\overline{x})$. Описанная процедура продолжается до первого нарушения неравенства вида (10).

3.2.2. Пусть неравенство вида (10) нарушено для полиномиальной регрессии, содержащей {линейные члены} $\cup \{M_{\mathbf{1}},\ M_2^1,\ \ldots, M_{t-k}^1\}$. Тогда конструируем полиномиальные регрессии из класса {линейные члены} $\cup \{M_{\mathbf{1}}\} \cup \{M_2^1\} \cup \ldots \cup \{M_{t-k-1}^1\}$, включая в них все возможные комбинации элементов $M_{t-k}^1$ (в первую очередь, исключая минимальные по величине члены), и для каждой комбинации проверяем выполнение неравенства вида (10). В случае его выполнения соответствующие элементы из $M_{t-k}^1$ исключаются. После проверки всех комбинаций элементов из $M_{t-k}^1$ описанная процедура повторяется для множеств $M_{t-k+1}^1$ и т. д. Вычислительная процедура выполнена либо когда окончилось время, выделенное для работы алгоритма, либо когда в результате перебора всех комбинаций исследуемое текущее множество $M_p^1$ осталось без изменений.

4. Пусть $f(\overline{x})$ – это результирующее выражение для многомерной полиномиальной регрессии. Тогда, считая все ее нелинейные члены известными, методом наименьших квадратов для многомерной линейной регрессии повторно оцениваем коэффициенты при ее линейных членах.

5. В случае, когда для проверочных данных возможен активный эксперимент в области $\forall_i\, |x_i| > 1$, считая все нелинейные члены известными, п.п. 1–3 реализуем только для линейных членов многомерной регрессии, оценки коэффициентов в которых были получены в п. 4.

**Выводы.** В статье изложен метод нахождения структуры и оценки коэффициентов исследуемой многомерной полиномиальной регрессии, заданной избыточным описанием с использованием активного эксперимента. Метод основан на:

1) свойствах нормированных ортогональных полиномов Форсайта [3, 4];

2) возможности сведения получения оценок коэффициентов многомерной полиномиальной регрессии, заданной избыточным описанием, к последовательному построению одномерных полиномиальных регрессий и соответствующих систем линейных уравнений [1–9] и метода исключения избыточных членов, основанного на основных идеях метода группового учета аргументов, а также кластерного анализа.

**Список литературы**


1. Zgurovsky M. Z., Pavlov A. A. *Combinatorial Optimization Problems in Planning and Decision Making: Theory and Applications*. Cham (Switzerland): Springer, 2019. 526 p. doi: 10.1007/978-3-319-98977-8
2. Згуровский М. З., Павлов А. А. *Труднорешаемые задачи комбинаторной оптимизации в планировании и принятии решений*. Київ: Наук. думка, 2016. 716 с.
3. Павлов А. А., Калашник В. В. Рекомендации по выбору зоны проведения активного эксперимента для одномерного полиномиального регрессионного анализа. *Вісник НТУУ "КПІ". Серія «Інформатика, управління та обчислювальна техніка»*. Київ: "ВЕК+", 2014. № 60. С. 41–45.

4. Павлов А. А., Калашник В. В., Коваленко Д. А. Построение многомерной полиномиальной регрессии. Регрессия с повторяющимися аргументами во входных данных. *Вісник НТУУ "КПІ". Серія «Інформатика, управління та обчислювальна техніка»*. Київ: "ВЕК+", 2015. № 62. С. 57–61
5. Худсон Д. *Статистика для физиков: Лекции по теории вероятностей и элементарной статистике*. 2-е изд. Москва: Мир, 1970. 296 с.
6. Згуровский М. З., Павлов А. А. *Принятие решений в сетевых системах с ограниченными ресурсами*. Київ: Наук. думка, 2010. 573 с.
7. Павлов А. А., Чеховский А. В. Построение многомерной полиномиальной регрессии. Активный эксперимент. *Системні дослідження та інформаційні технології*, 2009, № 1. С. 87–99
8. Павлов А. А., Чеховский А. В. Построение многомерной полиномиальной регрессии. Активный эксперимент с ограничениями. *Вестник Нац. техн. ун-та «ХПИ»: сб. науч. тр. Темат. вып.: Системный анализ, управление и информационные технологии*. Харьков: НТУ «ХПИ», 2009. № 4. С. 174–186.
9. Павлов А.А., Чеховский А.В. Сведение задачи построения многомерной регрессии к последовательности одномерных задач. *Вісник НТУУ "КПІ". Серія «Інформатика, управління та обчислювальна техніка»*. Київ: "ВЕК+", 2008. № 48. С. 111–112.
10. *Ивахненко А.Г.* Моделирование сложных систем. Киев: Вища школа, 1987.
11. Настенко Е., Павлов В., Бойко Г., Носовец О. Многокритериальный алгоритм шаговой регрессии. *Біомедична інженерія і технологія*, 2020. № 3, С. 48–53. doi: 10.20535/2617-8974.2020.3.195661
12. Коваленко Д. А. Застосування принципів багатовимірної поліноміальної регресії для розвідувального аналізу даних та знаходження лінії регресії. *Науковий огляд*, 2018. Том 3, № 46, С. 81–94.

**References (transliterated)**

1. Zgurovsky M. Z., Pavlov A. A. *Combinatorial Optimization Problems in Planning and Decision Making: Theory and Applications*. Cham: Springer Publ., 2019. 526 p. doi: 10.1007/978-3-319-98977-8
2. Zgurovsky M. Z., Pavlov A. A. *Trudnoreshaemye zadachi kombinatornoj optimizacii v planirovanii i prinjatii reshenij* [Intractable combinatorial optimization problems in planning and decision making]. Kiev, Nauk. dumka Publ. 716 p.
3. Pavlov A. A., Kalashnik V. V. Rekomendacii po vyboru zony provedenija aktivnogo jeksperimenta dlja odnomernogo polinomial'nogo regressionnogo analiza [Recommendations for choosing the zone of an active experiment for one-dimensional polynomial regression analysis]. *Visnyk NTUU "KPI". Seriya «Informatyka, upravlinnya ta obchislyuvalna tekhnyka»*. Kiev, "Vek+" Publ., 2014, no. 60, pp. 41–45
4. Pavlov A. A., Kalashnik V. V., Kovalenko D. A. Postroenie mnogomernoj polinomial'noj regressii. Regressija s povtorjajushhimisja argumentami vo vhodnyh dannyh [Multidimensional polynomial regression construction. Regression with duplicate arguments in the input]. *Visnyk NTUU "KPI". Seriya «Informatyka, upravlinnya ta obchislyuvalna tekhnyka»*. Kiev, "Vek+" Publ., 2015, no. 62, pp. 57–61
5. Hudson D.J. Statistics Lectures, Volume 2: Maximum Likelihood and Least Squares Theory. CERN Reports 64(18). Geneva, CERN, 1964. (Russ. ed.: Hudson D. *Statistika dlja fizikov: Lekcii po teorii verojatnostej i jelementarnoj statistike*. Moscow, Mir Publ., 1970. 296 p.). doi: 10.5170/CERN-1964-018
6. Zgurovsky M. Z., Pavlov A. A. *Prinyatie resheniy v setevykh sistemakh s ogranichennymi resursami* [Decision making in network systems with limited resources]. Kiev, Nauk. dumka Publ., 2010, 573 p.
7. Pavlov A. A., Chekhovskiy A. V. Postroenie mnogomernoj polinomial'noj regressii. Aktivnyj jeksperiment [Multidimensional polynomial regression construction. Active experiment]. *System research and information technologies*, 2009, no. 1, pp. 87–99
8. Pavlov A. A.., Chekhovskiy A. V. Postroenie mnogomernoj polinomial'noj regressii. Aktivnyj jeksperiment s ogranichenijami [Multidimensional polynomial regression construction. Active experiment with limitations]. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2009, no. 4, pp. 174–186
9. Pavlov A. A., Chekhovskiy A. V. Svedenie zadachi postroenija mnogomernoj regressii k posledovatel'nosti odnomernyh zadach [Reducing the problem of multivariate regression constructing to a sequence of one-dimensional problems]. *Visnyk NTUU "KPI". Seriya «Informatyka, upravlinnya ta obchislyuvalna tekhnyka»*. Kiev, "Vek+" Publ., 2008, no. 48, pp. 111–112
10. Ivahnenko A.G. Modelirovanie slozhnyh sistem. Kiev, Vyshha shkola Publ., 1987
11. Nastenko E., Pavlov V., Boyko G., Nosovets O. Mnogokriterial'nyj algoritm shagovoj regressii. *Biomedychna inzheneriya i tekhnolohiya* [Biomedical ingeneering and technology]. 2020, no. 3, pp. 48–53. doi: 10.20535/2617-8974.2020.3.195661
12. Kovalenko D. A. Zastosuvannya pryntsypiv bahatovymirnoyi polinomial'noyi rehresiyi dlya rozviduval'noho analizu danykh ta znakhodzhennya liniyi rehresiyi [Applying multiple polynomial regression principles for exploratory data analysis and regression analysis]. *Scientific Review*, vol. 3, no. 46, 2018, pp. 81–94.


*Поступила (received) 11.05.2020*

*Відомості про авторів / Сведения об авторах / About the Authors*


***Павлов Олександр Анатолійович*** – доктор технічних наук, професор, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», завідувач кафедри Автоматизованих систем обробки інформації та управління; м. Київ, Україна; ORCID: https://orcid.org/0000-0002-6524-6410; e-mail: pavlov.fiot@gmail.com

***Головченко Максим Миколайович*** – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», асистент кафедри Автоматизованих систем обробки інформації та управління; м. Київ, Україна; ORCID: https://orcid.org/0000-0002-9575-8046; e-mail: ma4ete25@ukr.net

***Павлов Александр Анатольевич*** – доктор технических наук, профессор, Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», заведующий кафедрой Автоматизированных систем обработки информации и управления; г. Киев, Украина; ORCID: https://orcid.org/0000-0002-6524-6410; e-mail: pavlov.fiot@gmail.com

***Головченко Максим Николаевич*** – Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», ассистент кафедры Автоматизированных систем обработки информации и управления; г. Киев, Украина; ORCID: https://orcid.org/0000-0002-9575-8046; e-mail: ma4ete25@ukr.net

***Pavlov Alexander Anatolievich*** – Doctor of Technical Sciences, Professor, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Head of Department of Computer-Aided Management and Data Processing Systems; Kyiv, Ukraine; ORCID: https://orcid.org/0000-0002-6524-6410; e-mail: pavlov.fiot@gmail.com

***Holovchenko Maxim Nikolaevich*** – National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Assistant of Department of Computer-Aided Management and Data Processing Systems; Kyiv, Ukraine; ORCID: https://orcid.org/0000-0002-9575-8046; e-mail: ma4ete25@ukr.net